

équivalences

La traduction médicale



Numéros 33/1-2
2006

Dynamique discursive et traduction des signes abrégatifs en biomédecine

Sylvie Vandaele
professeure agrégée
Manon Pageau

*Département de linguistique et de traduction
Université de Montréal*

1. Introduction

Un problème des plus délicats rencontrés dans la pratique de la traduction biomédicale est celui des signes abrégatifs, autrement dit abrégements, abréviations, sigles, acronymes et symboles. La biologie cellulaire et moléculaire, ainsi que la génétique, sont des spécialités particulièrement touchées par leur prolifération incontrôlable, car ces domaines de recherche fort dynamiques découvrent ou fabriquent une quantité importante de molécules, de cellules ou de microorganismes, ce qui entraîne une activité frénétique de dénomination. Il s'ensuit que les traducteurs médicaux et scientifiques se heurtent fréquemment au traitement de ces formes linguistiques. Toutefois, la traduction humaine n'est pas la seule concernée : les signes abrégatifs constituent aussi un problème de fond en traitement automatique de la langue (TAL), qu'il s'agisse d'extraction de connaissances, d'indexation ou de production de résumés automatiques, notamment en raison des problèmes de désambiguïsation soulevés par leur usage (Pustejovsky 2002).

Les difficultés de traduction sont de plusieurs ordres. Tout d'abord, la prolifération des signes abrégatifs entraîne un obscurcissement considérable du message pour le non-spécialiste :

- (1) *Bien que la co-expression du CRLR avec RAMP1 résulte dans l'expression d'un récepteur du CGRP, sa co-expression avec RAMP2 ou RAMP3 donne naissance à un récepteur qui a les propriétés pharmacologiques d'un récepteur de l'adrénoméduline [...]. (Bouvier et al. 2000 : 649)*
- (2) *CAP4 was recently identified as a novel cysteine protease of the ICE/Ced-3 family that contains a prodomain homologous to FADD, and was designated MACH or FLICE*

[...]. *CAP3 has been identified as the prodomain of FLICE, and is presumably generated during CD95-induced proteolytic activation of FLICE/MACH [...].* (Pelengaris 1997 : 474)

On constate d'emblée que l'absence de motivation apparente ne facilite pas la compréhension du texte – et *ipso facto*, sa traduction. La forme développée, sensée désambigüiser, n'est pas toujours indiquée dans le texte. Certes, certains signes abrégatifs sont assez connus des lecteurs potentiels pour que leur signification ne soit pas systématiquement indiquée en contexte (par ex., *DNA* et *ADN*¹). Cependant, dans les domaines de pointe, le nombre des « initiés » peut parfois être très restreint. Par contre, Internet constitue une aide appréciable en matière de désambigüisation, car différents comités internationaux tiennent à jour, en permanence, les nomenclatures des organismes vivants et des molécules utilisées en recherche et les rendent accessibles en ligne².

Autre difficulté, un même signe abrégatif peut être associé à différents référents : le signe abrégatif *Ras* ou *ras* sera utilisé pour le gène (*le gène ras*) et la protéine codée par celui-ci (*Ras, la protéine Ras*) ou la famille de protéines apparentées (*la famille Ras*), ou encore, pour une voie de communication cellulaire mettant en jeu ladite protéine (*la voie Ras*). Dans certains cas, des souches de cellules ou de souris utilisées dans une recherche particulière exprimant ou non une molécule donnée adoptent le nom de cette dernière (*les cellules Smad4+/-* : des cellules possédant un seul allèle (+/-) du gène codant pour la protéine *Smad4*). Par conséquent, la consultation des bases de données peut se révéler utile voire indispensable, mais là encore, l'analyse du contexte constituera une étape essentielle pour le processus de traduction.

Vient ensuite la question de l'équivalence : la grande majorité des formes apparaissant pour la première fois par l'intermédiaire de publications en anglais – *lingua franca* de la biomédecine, comme de bien d'autres domaines – se pose la question de savoir s'il existe un équivalent dans la langue d'arrivée, le français dans notre cas. Ainsi, certaines formes anglaises (*DNA*) sont francisées (*ADN*), d'autres, non (*VLDL, PDGF*). Bien que certains auteurs aient fortement argumenté en faveur de la francisation des signes abrégatifs (Jammal 1992), il apparaît clairement que le nombre de formes non francisées est largement majoritaire. De nombreux ouvrages (pour n'en citer que quelques uns : Maillet 1996 ; Touati

1 Les signes abrégatifs spécifiques au français sont en caractère gras. Dans les exemples, les italiques du texte original ont été converties en caractère droit.

2 Une liste des principaux sites de nomenclatures est accessible à la page <http://www.mapageweb.umontreal.ca/vandaels>, onglet « Liens ».

1994 pour le français ; Jablonski 2005 pour l'anglais ; Dionne et Boulay 1998 pour les équivalences anglais-français), ou des ressources en ligne tentent de colliger le plus grand nombre de signes abrégatifs possibles, mais l'équivalence, lorsqu'elle existe, n'est pas toujours mentionnée. Enfin, la vitesse à laquelle ces signes se multiplient rend la consultation de ces sources assez souvent infructueuse (notamment pour les formes les plus récentes) et impose la recherche de contextes *ad hoc*. Ainsi, le traducteur devra rechercher les nouvelles formes dans les articles scientifiques les plus récents, par exemple en ayant recours à Medline ou à Current Content (Vandaele 2001) et décider par lui-même quelle stratégie adopter.

Enfin, le comportement discursif, ou syntagmatique, dans la langue d'arrivée est problématique. Si le français a établi des règles générales d'usage relatives aux abréviations et aux symboles³, leur application en biomédecine semble, à première vue, erratique et incertaine : les formes développées sont souvent absentes, le genre est parfois difficile à déterminer et l'absence d'équivalents français, que ce soit pour le signe abrégatif ou la forme développée, semble fréquente. De plus, les signes abrégatifs se retrouvent parfois en apposition d'un nom commun (*le gène COL7A1*) ou seuls, avec un déterminant (*l'AMPc*, *l'ADN*) ou sans (*L'association d'un complexe contenant Smad à la SBS de COL7A1 requiert sa phosphorylation* (Mauviel 1999)). Se pose ainsi la question de savoir dans quelle mesure ces formes sont autonomes, s'il est possible de dégager certaines régularités permettant d'en comprendre le fonctionnement et, partant, de comprendre comment les traiter dans le processus de traduction.

Nous nous proposons, dans le présent article, d'explorer la question grâce à une recherche effectuée dans un corpus de biologie cellulaire constitué d'articles de synthèse⁴ rédigés en français. L'ensemble des signes abrégatifs a été relevé afin d'en analyser la nature et le comportement spécifiques. À la lumière de cette analyse, nous proposerons des pistes de solutions visant à aider à la prise de décision en traduction. Nous ne prétendons pas à l'exhaustivité, mais

3 Les ouvrages de référence habituels donnent des règles générales. Ainsi, «lors d'une première mention dans un texte, il importe d'écrire au long la signification de toute abréviation non usuelle, tout sigle, acronyme ou symbole non courant» (De Villers 2001). Les accents et traits d'union appartenant au mot ou au syntagme abrégés sont conservés, et le point abrégatif des sigles ou des abréviations est fusionné avec le point final de la phrase. En ce qui concerne le comportement syntaxique, aucune de ces formes linguistiques ne prend la marque du pluriel. De Villers indique que les sigles, comme les acronymes, «sont du genre et du nombre du mot principal de la désignation abrégée.»

4 Ce type d'articles vise à faire le point sur une question, par opposition aux articles de recherche, qui présentent des résultats originaux (avec les sections méthodologie, résultats, discussion), et aux articles de vulgarisation, qui transmettent l'information au grand public.

espérons que les paramètres d'analyse utilisés permettront d'étendre la réflexion à l'ensemble des phénomènes possibles, non seulement en biomédecine, mais aussi dans d'autres domaines.

2. Quelques clarifications terminologiques

De façon générale, les différents auteurs ayant abordé la question ne s'accordent pas sur la nature exacte des signes abrégatifs et leur classification (par ex., Calvet 1980 ; Kocourek 1982 ; Germain 1988 ; Mortureux 1994 ; Monnier 1994), non plus que les dictionnaires (De Villers 2001 ; Robert 2001), ni les grammaires (Grevisse 1993 : §§110 et 189). Germain (1988) souligne que le terme *abréviation*, notamment, est utilisé par différents auteurs dans un sens large ou restreint. Pour pallier cet inconvénient, nous lui avons emprunté la notion générale de signe abrégatif, pour distinguer ensuite les abrégements, les abréviations, les sigles et les acronymes. Tandis que pour Grevisse, les acronymes semblent constituer un type particulier de sigles, De Villers (2001) les distingue d'emblée. Nous avons retenu, pour les distinguer, le critère de la prononciation alphabétique ou syllabique (Germain, 1988) :

- **signe abrégatif** : de façon générale, toute forme réduite graphiquement, mettant en jeu des parties de termes simples ou complexes. Ceci exclut formellement la réduction terminologique⁵ (Collet 2000).
- **abrégement** : signe abrégatif résultant de l'amputation d'une ou de plusieurs syllabes (*métro*, pour *métropolitain* ; *glu*, pour *glutamine*).
- **abréviation** : signe abrégatif résultant de la suppression de lettres, avec le maintien d'au moins l'initiale (*km*, pour *kilomètre* ; *vx*, pour *vieux* ; *mM*, pour *millimolaire*), ne donnant lieu ni à une prononciation syllabique, ni à une prononciation alphabétique.
- **sigle** : signe abrégatif résultant de l'accolement de lettres et/ou de groupes de lettres, généralement initiales ou finales, auxquels peuvent s'ajouter des chiffres et/ou des lettres grecques, et donnant lieu à une prononciation alphabétique (*NIH-3T3*, *CHO*, *RCPG*).
- **acronyme** : signe abrégatif résultant de l'accolement de lettres et/ou de groupes de lettres initiales ou finales dont la prononciation est syllabique (*CEGEP*, *sida*).

Le symbole, quant à lui, se caractérise par son caractère conventionnel translinguistique. Sa définition a été empruntée à De Villers (2001) :

⁵ La réduction terminologique consiste à employer, en discours, la tête d'un syntagme terminologique complexe, qui prend alors une valeur de reprise anaphorique.

- **symbole** : « signe conventionnel constitué par une lettre (*h*, heure), un groupe de lettres (*km*, kilomètre), un groupe de lettres et de chiffres (H_2O , symbole chimique), un signe (\$, dollar), un pictogramme (un schéma de punaise (*pin*) indiquant une note). Le symbole, indépendamment des frontières linguistiques, sert à désigner de façon très concise, un être, une chose, une grandeur, une réalité. [Ils] s'emploient principalement dans les domaines scientifiques et techniques : symboles chimiques, mathématiques, symboles des unités monétaires, des unités de mesure ».

Nous verrons que nous serons amenées à distinguer, dans le présent travail, les symboles courants des symboles dénominatifs spécifiques à la biomédecine qui sont, à l'origine, des sigles, des abrègements ou des acronymes issus de travaux de recherche, mais qui sont ensuite régulés par des nomenclatures internationales, acquérant ainsi un statut translinguistique.

3. Méthodologie

3.1. Corpus

Afin d'assurer la cohérence thématique du corpus, les textes retenus, rédigés en français et non traduits, devaient traiter de communication cellulaire et plus spécifiquement du rôle joué par les protéines. Ce choix a été fait en raison de leur rôle essentiel dans les mécanismes cellulaires et de leurs nombreuses dénominations. Il semblait aussi, dans les observations préliminaires, que leurs dénominations étaient les plus problématiques. Bien que de nombreuses entités autres que les protéines soient évoquées dans ces textes, il faudra toutefois garder en tête que ce choix introduit un biais dans les valeurs chiffrées, dans la mesure où des abréviations correspondant à des noms de souches cellulaires, de microorganismes ou à des substances chimiques pourront être, par comparaison, sous-représentées. Par conséquent, certains types de sigles abrégatifs ont pu nous échapper.

Par ailleurs, les auteurs devaient être des chercheurs spécialisés dans le domaine, de même que le lectorat cible. Ont été exclus les textes journalistiques ou de vulgarisation et les textes traduits, ainsi que les pages Web, même rédigées par des chercheurs, en raison de l'absence généralisée de révision et d'évaluation. La seule revue internationale scientifique de haut niveau francophone publiant uniquement des articles de synthèse dans le domaine biomédical, *m/s médecine-sciences*, remplissait toutes les conditions. Cette revue spécialisée vise à faire la synthèse des recherches fondamentales en cours sur les mécanismes physiopathologiques aux plans cellulaires et moléculaires. Son

public cible est constitué par les chercheurs en biomédecine, elle est internationale et les articles sont soumis à une triple évaluation anonyme par les pairs (Bergeron et Kahn 1995 : 315). Les standards langagiers du domaine sont donc fortement influencés par les articles qui y sont publiés.

Le corpus compte 64 000 occurrences⁶ provenant de 36 articles écrits entre 1992 et 2001, avec une moyenne de 1778 occurrences par article. Bien que le corpus puisse paraître restreint, l'échantillonnage des auteurs est diversifié : les textes proviennent de 35 auteurs ou équipes de recherche différents. Comme le nombre de chercheurs francophones spécialisés est lui-même petit, ces auteurs, en raison du statut particulier de la revue et de son audience, jouent un rôle clé dans l'établissement de l'usage. Par conséquent, le corpus ainsi constitué peut être considéré comme un échantillon représentatif des standards en usage dans le domaine.

3.2. Recueil des données

Tous les articles ont été numérisés, puis soumis à une reconnaissance de caractères à l'aide du logiciel *OmniPage v.8* (sous Mac OS 9)⁷ et convertis en format électronique. Les signes abrégatifs (apparaissant au moins une fois) ont été repérés manuellement comme étant toute suite de lettres (en majuscules ou en minuscules) ne constituant pas un mot de la langue, comportant ou non des exposants ou des indices, des chiffres ou des lettres grecques.

Ont été également repérées les formes développées et les dénominations communes correspondantes éventuelles. La recherche de ces formes et des contextes définitoires ou explicatifs a été réalisée à l'aide d'un concordancier, *Le Concordeur 3.0* (David Rand, Université de Montréal) au sein même du corpus. Des sources externes ont été consultées lorsque le corpus ne contenait pas l'information recherchée⁸ : différents dictionnaires de langue générale en anglais et en français, dictionnaires bilingues spécialisés, dictionnaire de médecine et de biologie, bases de données terminologiques et bibliographiques. Les données recueillies ont été colligées dans une base de données informatisée créée à l'aide du logiciel *FileMaker Pro* (v. 5 à 7 ; sous *Mac OS9* puis *Mac OSX*).

6 Occurrence (*token*) : pris dans le sens de chaîne de caractères isolée par une espace ou une marque de ponctuation.

7 Les articles de *m/s médecine-sciences* sont maintenant accessibles en format PDF sur Internet. Archives depuis 1996 : <http://ist.inserm.fr/basismedsci/msu.html> ; Numéros récents : <http://www.erudit.org/revue/ms/>.

8 Voir Pageau, M. (2003) pour une description détaillée des nombreuses ressources utilisées, ainsi que Vandaele (2001a) pour une bibliographie d'ouvrages médicaux et Vandaele (2001b) pour l'utilisation des bases de données bibliographiques telles que Medline et l'INIST.

4. Typologie et morphologie

Sur l'ensemble du corpus, 701 signes abrégatifs et symboles différents ont été relevés. Les symboles courants ayant un statut quelque peu à part en raison de leur valeur translinguistique bien établie, ils seront traités séparément. Les autres signes seront présentés à l'aide de la typologie annoncée à la section 2, à laquelle quelques modifications ont dû être apportées.

4.1. Symboles courants

Différents symboles courants en biomédecine ont été relevés. On rencontre, comme il est naturel en sciences, des unités de mesure (*mM* [millimole] ; *nm* [nanomètre] ; *kb* [kilobase]), des expressions mathématiques (K_d , pour la constante de dissociation d'un ligand à l'égard de son récepteur ; K_m , pour la constante de Michaelis), des formules chimiques (*NO*, qui est la formule chimique du monoxyde d'azote) et des symboles désignant les phases du cycle cellulaire (*S*, *M*, *G0*, *G1*, *G2*).

Un cas particulier est représenté par les lettres symbolisant les acides aminés et les nucléotides. Dans le cas des acides aminés (au nombre de vingt)⁹, on relève la coexistence, d'une part, de symboles constitués le plus souvent par la première lettre ou l'une des lettres du nom des acides aminés (sauf pour l'acide aspartique [D], la phénylalanine [F], la lysine [K], la glutamine [Q] et le tryptophane [W]) et, d'autre part, par des abrégements formés généralement par les trois premières lettres du nom de l'acide aminé (sauf pour l'isoleucine [Ile], l'asparagine [Asn], la glutamine [Gln] et le tryptophane [Trp]). En contexte, l'acide aminé est généralement désigné par son nom complet :

(3) *L'adipsine, qui est une protéase à sérine, [...].* (Ailhaud 1998 : 858)

(4) *Ces motifs, appelés PEST car riches en proline (P), acide glutamique (E), acide aspartique (D), sérine (S) et thréonine (T), sont bordés par des régions basiques.* (Carillo 1995 : 731)

(5) *Parmi ces sites, la sérine 64 et la thréonine 45 pourraient participer à la liaison à eIF-4E. Ces phosphorylations sont, semble-t-il, effectuées par la kinase mTOR.* (Rousseau 2001 : 337)

L'exemple (5) fait état d'une particularité : dans le domaine de connaissances considéré, les acides aminés font souvent partie intégrante de protéines.

9 Alanine : A, Ala ; cystéine : C, Cys ; acide aspartique : D, Asp ; glutamine : E, Glu ; phénylalanine : F, Phe ; glycine : G, Gly ; histidine : H, His ; isoleucine : I, Ile ; lysine : K, Lys ; leucine : L, Leu ; méthionine : M, Met ; asparagine : N, Asn ; proline : P, Pro ; glutamine : Q, Gln ; arginine : R, Arg ; sérine : S, Ser ; thréonine : T, Thr ; valine : V, Val ; tryptophane : W, Trp ; tyrosine : Y, Tyr. À noter que seules certaines formes ont été relevées au sein du corpus.

Leur nom est ainsi associé à un chiffre qui indique leur position dans la chaîne protéique : *sérine 64*, ce qui indique qu'en 64^e position se trouve un résidu sérine. C'est cette même idée d'acide aminé intégré dans la chaîne protéique qu'exprime l'expression *résidu + [nom d'acide aminé]*, comme dans l'exemple suivant :

(6) *Le reste de la molécule est formé d'une région riche en résidus proline et sérine contenant plusieurs résidus tyrosine phosphorylables.* (Lecoq-Lafon 1999 : 69)

Pour la même raison, l'abréviation de trois lettres est également retrouvée suivie d'un chiffre, mais au sein d'une phrase elle sera précédée du terme *résidu* :

(7) *Cinq sites de phosphorylation (Ser 64, 82 et Thr 36, 45, 69) semblent être impliqués dans la régulation de l'activité de 4E-BP1.* (Rousseau 2001 : 336)

(8) *Le site de phosphorylation d'eIF-4E est le résidu Ser 209.* (Rousseau 2001 : 336)

On retiendra que les formulations suivantes, qui ne sont pas retrouvées dans le corpus, sont incorrectes : **la Ser est un acide aminé qui...* ; **Ser est un acide aminé qui...*

Quant au symbole constitué d'une seule lettre, il est parfois utilisé isolément de façon ponctuelle et à titre informatif (exemple 4), mais le plus souvent dans des suites dénotant des séquences d'acides aminés constituant une chaîne peptidique :

(9) *Ce site, constitué de deux tyrosines phosphorylables (séquence YpVHVNATYpVNV) fixe et active plusieurs molécules.* (Lecoq-Lafon 1999 : 69)

Le même comportement est observé pour les symboles d'une lettre correspondant aux noms de nucléotides, tels que *A* (adénine), *U* (uracile), *G* (guanine), *C* (cytosine), *T* (thymine) : ces symboles apparaissent dans des suites de lettres dénotant des séquences nucléotidiques :

(10) *L'initiation de la synthèse protéique requiert la reconnaissance du codon d'initiation AUG d'un ARNm par un ribosome.* (Poulin 2000 : 617)

Bien que le détail des formes anglaises ne soit pas abordé dans le présent travail, précisons que le comportement en anglais des symboles est sensiblement le même, aux particularités syntaxiques et typographiques près (espaces, par ex.).

4.2. Signes abrégatifs : abrègements, acronymes et sigles

À part les symboles courants mentionnés ci-dessus, 626 signes abrégatifs différents ont été relevés. Le corpus comptant 9186 types d'occurrences différentes (non lemmatisées), 6,8 % des formes sont donc des signes abrégatifs. Chaque signe apparaissant au moins une fois (le plus fréquent est *GTP*, 70 occur-

rences), la fréquence des signes abrégatifs dans l'ensemble du corpus (64 000 occurrences au total) est supérieure à 1 %. À titre de comparaison, le terme le plus fréquent est *protéines* (au 17^e rang, 723 occurrences non lemmatisées). Ces quelques chiffres soulignent la fréquence élevée de ces formes et confirment l'importance de savoir maîtriser leur traitement.

Certaines des formes relevées appartiennent sans équivoque possible à l'une ou l'autre des catégories retenues : nous avons ainsi relevé dans le corpus des abrègements (*Ub*, pour *ubiquitin*), des sigles (**ADN**, pour *acide désoxyribonucléique* ; **CBC**, pour *cullin 2-elongin B-elongin C*), des acronymes (**sida**, pour *syndrome d'immunodéficience acquise*). Le tableau 1 présente la répartition de chacune des catégories. Plutôt que de créer des catégories mixtes, nous avons plutôt, pour quelques cas complexes, priorisé certains paramètres :

- Les signes abrégatifs simplement suivis d'un chiffre ont été classés dans la même catégorie que la forme sans chiffre : par exemple, *Cul1* [sic] (pour *Cullin-1*) est classé dans les abrègements, *SV40* (pour *simian virus 40*) dans les sigles, *Tap1* (pour *Transport Associated Antigen-1*) dans les acronymes. La raison de ce choix est que le chiffre ne fait qu'indiquer une espèce au sein d'une famille.
- La prononciation anglaise a, dans la mesure du possible, été prise en compte pour le classement parmi les acronymes : c'est en effet une norme du domaine partagée internationalement. Il est cependant possible qu'une variation de la prononciation puisse exister au sein des locuteurs.
- Quelques signes abrégatifs possédaient une structure complexe (par ex. *Ship*, pour *SH2 containing inositol phosphatase*), mais ils ont été comptabilisés avec les acronymes. Toutefois, quelques-uns, qui étaient du type *t-SNARE*, ont été comptabilisés parmi les sigles complexes, en raison de la lettre *t* qui est le signe abrégatif pour *target*.

Abrègements	11	1,75%
Acronymes	63	10%
simples	59	
complexes (mixtes)	4	
Sigles	552	88,70%
simples	376	68,06%
complexes	96	15,30%
<i>juxtaposés</i>	32	5,11%
<i>mixtes</i>	35	5,59%
<i>de motivation non déterminée</i>	29	4,29%
<i>de motivation indéterminée</i>	80	12,77%

Tableau 1 – Répartition des signes abrégatifs

Les sigles comportent toujours plus d'une lettre, souvent des chiffres (*M6P*) ou des symboles grecs (*Mato2*). La grande majorité d'entre eux sont simples (au moins 376 sur 552, soit environ 68 %), c'est-à-dire qu'ils sont directement motivés par la forme développée anglaise ou française (EGF, pour *epidermal growth factor*). Une petite centaine de sigles présentent une morphologie complexe. Deux catégories de sigles ont été distinguées : les sigles juxtaposés et les sigles complexes. Les sigles juxtaposés sont formés par l'association d'au moins deux signes abrégatifs simples : ils peuvent être continus (*UbcH5*), séparés par un trait d'union (*KDEL-ERD2*) ou une barre oblique (*FRAP/mTOR*), ou encore disposés en exposant ou en indice (*APC/CCDC20*, *p59^{lyn}*).

Les sigles complexes comportent un ou plusieurs autres signes abrégatifs dans leur forme développée. Ainsi, la forme développée de *t-SNARE est target-Soluble NSF acceptor receptor*, *NSF* se décompose à son tour en *NEM sensible factor* et enfin, *NEM*, en *N-ethylmaleimide* ; le sigle *PDZ* est un amalgame des signes suivants : *PSD-95*, *Discs-large* et *ZO-1*. *ARF-GAP* se développe en *ADP-ribosylation factor – GTPase activating protein* et *FRAP/mTOR*, en *FKBP-rapamycin associated protein/mammalian target of rapamycin*.

Enfin, 80 sigles sont restés indéterminés, de même que 29 sigles complexes, faute d'avoir pu caractériser tous leurs composants.

Les abrègements sont les plus motivés, et sont aisément mémorisables : leurs formes développées elles-mêmes évoquent certaines propriétés du référent, et leur graphie est généralement francisée (*ubiquitin/ubiquitine*, *obese/obèse*, *envelop/enveloppe*...). La motivation des acronymes est moins directe, mais ils restent aisément mémorisables, certains d'entre eux ayant une consonance humoristique ou évocatrice (*Myc-Max*, *Sea*, *Ship*...). C'est le mode privilégié de dénomination des oncogènes (*Mos*, *Nef*, *Rab*, *Rap*, *Ras*, *Rev*, *Sea*, *Sim*, *Src*, *Abl*, *Myc*, *Max*...), qui sont des gènes codant pour des protéines spécifiquement impliquées dans processus tumoraux.

Les sigles complexes, mixtes ou juxtaposés, semblent être une caractéristique du domaine biomédical. Peu motivés, ces sigles semblent plutôt rares dans la langue commune. Il est toutefois probable qu'ils soient représentés dans d'autres domaines spécialisés, comme l'informatique. Monnier (1994) n'en a relevé que trois exemples dans le domaine de l'industrie spatiale. Germain (1988 : 65) mentionne « un cas plutôt inusité d'un sigle siglé dans *SECAL (Société d'électrolyse et de chimie Alcan)* où les deux dernières lettres appartiennent à un autre sigle, *ALCAN (Aluminium Canada)* ». Leur grand nombre en biomédecine reflète la complexité des mécanismes en jeu dans le fonctionnement des êtres vivants. En effet, bien que le néophyte ne s'y retrouve guère, il s'agit, pour les chercheurs, de véritables « noms de code », permettant une économie d'espace. Le fait qu'un certain nombre d'entre eux soient peu ou pas

motivés en apparence constitue un obstacle important à leur traitement, mais il est probable que leurs inventeurs ne les ont pas créés au hasard. L'examen de leurs propriétés linguistiques s'impose donc afin de rechercher des régularités permettant de tracer une ligne de conduite en matière de traduction.

5. Propriétés linguistiques

5.1. Classes référentielles représentées

Un paramètre important dans le comportement syntagmatique des signes abrégatifs est la classe à laquelle appartient le référent dénoté.

Catégorie référentielle	Signe abrégatif	Forme développée apparaissant dans le corpus
Gènes et ensemble de gènes	<i>CMH</i>	<i>complexe majeur d'histocompatibilité</i>
Complexe macromoléculaire	<i>APC/C</i>	<i>anaphase promoting complex/ cyclosome</i>
Macromolécules (acides nucléiques, protéines et peptides...)	<i>ADN ARN ASP CGRP LMP2</i>	<i>acide désoxyribonucléique acide ribonucléique acylation stimulating protein calcitonin gene-related peptide low molecular weight protein-2</i>
Molécules fonctionnelles (enzymes, facteurs, récepteurs...)	<i>PKA UCDEN EGF RCPG SNARE</i>	<i>protéine kinase A ubiquitin-conjugate degrading enzyme facteur de croissance épidermique récepteur couplé aux protéines G récepteur des SNAP</i>
Parties de macromolécules	<i>domaines DH, domaines Btk 5'UTR NLS</i>	<i>région 5' non traduite signal de localisation nucléaire</i>
Espèces chimiques diverses	<i>RLO</i>	<i>radicaux libres de l'oxygène</i>
Animaux (rat, souris)	<i>souris A/J</i>	
Microorganismes et agents pathogènes (virus, bactéries)	<i>VSV VEMC VIH</i>	<i>virus de la stomatite vésiculaire virus de l'encéphalomyocardite virus de l'immunodéficience humaine</i>
Cellules ou lignées cellulaires	<i>NIH3T3, CHO cellule HeLa</i>	
Compartiments cellulaires	<i>ER RTG VS</i>	<i>réticulum endoplasmique réseau trans-golgien vésicule sécrétoire</i>
Affections	<i>sida</i>	<i>syndrome d'immunodéficience acquise</i>

Techniques médicales	<i>RMN</i> <i>TEP</i> <i>AMP</i>	<i>résonance magnétique nucléaire</i> <i>tomographie par émission de positons</i> <i>assistance médicale à la procréation</i>
Paramètre	<i>DL50</i>	<i>dose létale 50</i>

Tableau 2 – Catégories référentielles relatives aux signes abrégatifs relevés en corpus

5.2. Apposition

Les signes abrégatifs sont fréquemment situés en apposition (plus de 300 occurrences dans le corpus). Dans certains cas, il s'agit de différencier le référent de signes abrégatifs « homonymes » : *le gène rho, la protéine Rho, la petite protéine G rho*. En effet, le signe, qui peut fonctionner de façon autonome, peut se révéler ambigü : *La relation entre rho, ARF et les protéines G trimériques n'a cependant pas encore été établie* (Pauloin 1993 : 924). En principe, le nom du gène et celui de la protéine sont différenciés par une marque typographique (majuscule/minuscule, ou bien italique/caractère droit), mais bien que des normes aient été définies par les comités chargées des nomenclatures¹⁰, les contraintes éditoriales ou simplement des fluctuations d'usage peuvent générer une certaine ambiguïté.

On observe une forme de redondance indirecte, déjà observée par Monnier (1994 : 105-106), lorsque le nom auquel est apposé le signe abrégatif est soit l'équivalent de la tête de la forme développée anglaise : *les kinases ERK (extracellular regulated kinase)* (Charron et Jeannotte 1999 : 1155), *l'hormone de croissance GH (GH pour growth hormone)*, soit la tête elle-même : *le virus VIH-1 (virus de l'immunodéficience humaine-1)*. On peut interpréter l'existence de ce type de redondance par le fait que les sigles pour lesquels ce phénomène se produit se comportent comme des formes indissociables, car ils ont en fait un statut de symbole, ce qui n'est pas le cas de *VIH (virus de l'immunodéficience humaine ; *virus VIH est une forme impropre)*. Aussi, de rares occurrences, du type la protéine *acylation stimulating protein (ASP)* (Ailhaud 1998 : 859), constituent une fluctuation par rapport à un usage relativement stable par ailleurs, qui consisterait plutôt à écrire : *la protéine ASP (acylation stimulating protein)*. Autre exception, dans le cas de *gp160*, le signe se révèle dissociable (*glycoprotéine 160*), alors que dans le cas de *p53*, il ne semble pas l'être (*protéine p53*). L'origine de cette différence de comportement n'est pas expliquée, étant donné que la structure est la même dans les deux cas (*protéine ou glycoprotéine + chiffre indiquant la valeur du poids moléculaire*).

¹⁰ «It is recommended that gene and allele symbols are underlined in manuscript and italicized in print ; protein symbols should be represented in standard fonts.» (Wain, 2002)

Enfin, le signe abrégatif peut être apposé à un nom dénotant autre chose que le référent auquel il correspond : il se comporte alors comme un adjectif relationnel, et à ce titre, la valeur du lien sémantique le reliant au nom est variable. Ainsi, *AA* est le sigle de *acide arachidonique*, et voie *AA* signifie « voie à laquelle participe l'acide arachidonique ». On repère, pour *TK* qui est le sigle de *tyrosine kinase*, diverses formes apposées : *l'activité TK* (activité enzymatique de type/manifestée par la tyrosine kinase), *le domaine TK* (domaine protéique ayant une activité tyrosine kinase), *un récepteur TK* (récepteur possédant un domaine ayant une activité tyrosine kinase).

5.3. « Homonymie » des signes abrégatifs et fluctuations de la graphie

Un certain nombre de « formes homonymiques » (Germain 1988 : 64) ont été relevées. Plusieurs sigles peuvent dénoter des référents différents, et donc plusieurs formes développées sont possibles, par exemple *AMP*, *SCF* et *Bcr*. Si la plupart des sigles se différencient à l'écrit par l'emploi de lettres majuscules et minuscules, par exemple *ARNt* et *Arnt*, *ASP* et *Asp*, *CDK* et *Cdk*, *PH* et *pH*, *TG* et *tG*, *THR* et *Thr*, certains ont la même graphie. Ainsi, *AMP* peut correspondre à *adénosine monophosphate* ou à *assistance médicale à la procréation*. *SCF* peut correspondre à *stem cell factor* ou à *Skp1-CDC53 (cullin)-F-box-protein*, dans les contextes suivants, pourtant tirés du même article (Gisselbrecht 1999 : 1263) : [...] *la réponse proliférative au stem cell factor (SCF)* et plus loin : *le complexe SCF pour Skp1-CDC53 (cullin)-F-box-protein*. *Bcr* peut correspondre à *breakpoint cluster region* ou à *B cell receptor*. Dans certains cas, la différence d'usage des majuscules et des minuscules est parfois liée à la convention qui devrait être appliquée (mais qui ne l'est pas toujours) pour rendre compte de la différence entre nom de gènes et nom de protéines (voir note 10).

Enfin, quelques cas de fluctuations graphiques ont été relevés, comme l'usage des tirets ou des espaces : *VIH1*, *VIH-1*, *VIH 1* réfèrent tous les trois au virus de l'immunodéficience humaine de type 1.

Au vu de cette analyse, la désambiguïsation peut être complexe si elle n'est pas faite *in situ* par l'auteur, imposant ainsi la consultation de dictionnaires et de nomenclatures, ainsi que des contre-vérifications dans d'autres articles.

5.4. Dérivation

Quelques signes abrégatifs donnent lieu à une dérivation nominale (*GTP* : *GTPase* ; *RNA* : *RNase*) ou adjectivale (*ATP* : *ATPasique*, *GABA* : *GABAergique* ; *GTP* : *GTPasique*). Le suffixe a une valeur sémantique : *-ase* indique que le référent ainsi nommé a une activité enzymatique dégradant, selon le cas, le GTP, l'ATP,

ou l'ARN. On remarquera l'effacement d'une voyelle dans ce dernier cas (**RNAase* est impropre). Quant à *-ergique* (*projection neuronale GABAergique*, *neurone GABAergique*), il indique que le neurone ou la projection neuronale fait usage de GABA (*acide gamma-aminobutyrique*) comme neurotransmetteur, ce suffixe se trouvant être l'équivalent français du suffixe *-ergic* anglais (*GABAergic*).

5.5. Usage du déterminant

Habituellement, un signe abrégatif adopte le comportement de la forme développée correspondante : dans ce cas, l'usage du déterminant est régulier, qu'il soit présent (*le VIH [virus de l'immunodéficience humaine]*, *l'ADN [acide désoxyribonucléique]*, *la PKA [protéine kinase A]*), ou absent en raison de la structure syntaxique commandée par la phrase (*[...] détectable par IRM [imagerie par résonance magnétique]*).

Toutefois, il est fréquent dans le corpus que les sigles soient employés sans déterminant (11, 12) ou qu'il y ait une alternance entre emploi et non-emploi du déterminant (13, 14).

(11) *eIF-4E appartient désormais à la grande famille des oncogènes. [...] eIF-4E semble donc être un marqueur de la transformation maligne...* (Rousseau 2001 : 336)

(12) *À ce jour, Btk est l'unique cas connu dans lequel des mutations dans le domaine PH sont à l'origine d'une maladie humaine.* (Broutin et Ducruix 2000 : 614)

(13) *Les ARF (ADP-ribosylation factor), identifiés à l'origine comme des cofacteurs [...], répondent aussi à la définition des petites protéines G.* (Pauloin 1993 : 920)

(14) *[...] Cette réaction d'échange GDP/GTP correspondant à l'activation d'ARF a lieu au niveau de la membrane car : (1) ARF nécessite des phospholipides pour subir cette réaction ; (2) ARF est myristoylé, et cet acide gras est nécessaire à son action [...].* (Pauloin 1993 : 920)

L'absence du déterminant suggère que le signe a acquis un statut de nom propre, parce qu'il sert à dénommer une instance de catégorie, c'est-à-dire une molécule particulière nommée ARF. C'est une situation rencontrée, effectivement, chaque fois que le sigle a un statut de symbole (c'est-à-dire qu'il a une valeur translinguistique) servant à identifier une instance moléculaire particulière. Kleiber (1996) démontre brillamment la différence entre nom propre et nom commun : le nom propre sert à dénommer l'instance de catégorie, tandis que le nom commun sert à dénommer la catégorie. Il est particulièrement intéressant de remarquer que le fonctionnement en tant que nom propre des signes abrégatifs renforce sans doute le mode de conceptualisation des molécules en tant qu'entité douées de capacité d'action (Vandaele et Lubin, 2005). Le fait que le sigle ne soit pas motivé renforce le caractère de nom propre du sigle. Les cas de fluctuation se rencontrent précisément lorsque le sigle est motivé par sa

forme développée, ce qui le « tire » du côté des noms communs, et plus encore lorsqu'il a fallu créer des instances supplémentaires de molécules en raison de découvertes nouvelles.

(15) [...] le signal riche en leucine de la protéine Rev du VIH-1 qui permet à certains ARN du virus d'être exportés vers le cytoplasme. (Dorseuil 1998 : 87)

(16) Tout d'abord, les protéines Env et Vpu codées par l'ARN de VIH1 contribuent à la rétention des molécules CD4. (Cosson et Letourneur 2000 : 637)

5.6. Genre

Ici encore, le comportement habituel d'un sigle est celui de la forme développée correspondante. Ainsi que l'explique Calvet (Calvet 1980 : 97) : « La langue a adopté la solution consistant à donner au sigle le genre du premier mot de sa source. Cette solution découle des structures syntaxiques du français selon lesquelles le substantif précède en général l'adjectif et le complément ». Cette règle est effectivement suivie par les sigles motivés par une forme développée possédant une syntaxe française, qu'ils se comportent comme des noms communs ou des noms propres. Par exemple, RCPG est masculin, comme récepteur (*récepteur couplé aux protéines G*) :

(17) Jusqu'à tout récemment, il était généralement admis que tous les événements de signalisation en aval d'un RCPG résultaient directement ou indirectement de l'activation d'une protéine G hétéro-trimérique. (Bouvier et al. 2000 : 647)

Si la forme développée est calquée de l'anglais, c'est également le genre de la tête du syntagme qui prévaut, compte tenu également de la syntaxe empruntée : le *GTP*, pour *guanosine triphosphate* (la tête *triphosphate* est masculin, tandis que *guanosine* est féminin).

Lorsque la forme développée est empruntée à l'anglais, c'est le genre de l'équivalent français de la tête du syntagme anglais qui prévaut : le *PGDF*, pour *platelet derived growth factor*, l'équivalent français donnant le genre au sigle étant *factor*. Enfin, lorsque le sigle est emprunté à l'anglais mais que la forme développée est traduite, le genre est également celui de la tête du syntagme français. Ainsi, *NLS* est au masculin, parce que la tête du syntagme en français est *signal*.

(18) Les protéines nucléaires, synthétisées dans le cytoplasme comme toutes les protéines, sont adressées au noyau grâce à un signal de localisation nucléaire (*NLS*, nuclear localisation signal) constitué d'une courte séquence d'acides aminés basiques, parfois bipartite. Le *NLS* est reconnu par des protéines spécifiques – dont les nucléoporines – et, peut-être par leur intermédiaire, se fixe aux pores de la membrane nucléaire. (Schreiber, de Murcia et Ménissier-de Murcia 1992 : 134)

Enfin, lorsqu'il n'y a pas de forme développée, c'est le genre du nom de la classe à laquelle appartient le référent qui donne le genre. Ainsi, les signes *myc*, *ras* etc. sont au masculin lorsqu'ils dénotent le gène, tandis les formes dénotant la protéine (*Ras*, *Myc*) sont au féminin.

Comme pour la question des déterminants, on relève quelques fluctuations. Dans un cas, le genre de la forme développée – en l'occurrence une dénomination commune de substance chimique, qui devrait suivre en principe le comportement décrit ci-dessus – n'est pas le même que celui du sigle, et ce dans le même article.

(19) [...] a cristallisé l'attention et l'inquiétude publiques autour du modèle moléculaire de la famille : **la** 2,3,7,8-tétrachlorodibenzo-p-dioxine (TCDD). [...] s'ajoutant à la sensibilité différente des espèces animales (de 1 à 5 000 pour **le** TCDD) [...]. (Lesca et Pineau 1999 : 1380)

Toutefois, une vérification de l'usage dans Internet a permis de constater qu'il s'agit d'un hapax et que le féminin est employé autant pour la forme développée que pour le sigle¹¹.

Une autre confusion a été observée pour le sigle *GTP* (*guanosine triphosphate*). C'est un cas typique de syntaxe empruntée à l'anglais, la tête étant *triphosphate* (masculin) et non *guanosine* (féminin). Tout comme pour les autres sigles désignant des nucléotides (ATP, ADP, AMP, etc.), *GTP* est au masculin. Une seule occurrence au féminin a été rencontrée dans un article scientifique extérieur au corpus, mais rédigé par l'un des auteurs représenté dans celui-ci :

(20) [...] et de protéines liant la guanosine diphosphate (GDP) et la guanosine triphosphate (GTP) (les ARF, pour Adenosine-diphosphate ribosylation factor) qui permettent aux coatomères de s'accrocher à la membrane de la vésicule. (Goud 1995 : 257)

Un cas particulièrement intéressant de fluctuation est représenté par l'exemple suivant, dans lequel il semble que les auteurs aient opté pour le genre féminin, qui est probablement lié au fait que **le** *NSF* est **une** protéine. Lorsque le genre peut être déterminé soit par la tête du syntagme, soit par le nom de la catégorie à laquelle appartient le référent, le genre de la tête du syntagme devrait l'emporter. Toutefois, dans le cas présent, le fait que la forme développée reste en anglais induit probablement une hésitation quant à la solution à adopter.

(21) Un criblage par la méthode de double hybrides en levure, a récemment révélé que la bêta-arrestine-1 peut aussi s'associer à une **protéine** impliquée dans les processus de fusion membranaire, **la** "N-ethyl-maleimide sensitive factor" (*NSF*). (Bouvier et al. 2000 : 647)

11 Voir entre autres le site Internet : ONTERM (Service de traduction du gouvernement 2003).

6. Équivalents des signes abrégatifs et des formes développées

Les 626 signes abrégatifs relevés ne se comportent pas tous de la même façon pour ce qui est de l'équivalence du signe lui-même, d'une part, et de sa forme développée, d'autre part. Le tableau 3 présente la répartition des différentes catégories de sigles.

Signes relevés	Formes développées françaises	Signes et formes développées anglaises ^a	Nombre
francisés ex. : <i>ADN</i>	canoniques acide désoxyribo-nucléique	<i>DNA</i> <i>desoxyribonucleic acid</i>	28 (4,47 %) ^b
francisés mais identiques à l'anglais ex. : <i>AA</i>	canoniques acide arachidonique	<i>AA</i> <i>arachidonic acid</i>	63 (10,06 %) ^a
identiques à l'anglais ex. : <i>ATP</i>	calquées adénosine triphosphate	<i>ATP</i> <i>adenosine triphosphate</i>	45 (7,18 %) ^a
empruntés ex. : <i>GH</i>	canoniques hormone somatotrope	<i>GH</i> <i>growth hormone</i>	157 (25,07 %) ^a
empruntés ex. : <i>IRPE</i>	non identifiées	<i>IRPE</i> <i>internal repositioning element</i>	333 (53,19 %) ^a

a Selon le cas, présents ou non dans le corpus, ou bien identifiés dans d'autres sources.

b Pourcentage calculé par rapport au nombre total de sigles relevés dans le corpus.

Tableau 3 – Répartition des différents types de signes abrégatifs du corpus

Sur les 626 signes abrégatifs relevés, 91 (14,53 %) ont une forme développée française canonique, c'est-à-dire que cette dernière respecte les contraintes de formation terminologiques et syntaxiques du français. Si 28 d'entre eux (4,47 %) correspondent exclusivement à une forme développée constituée de mots français, 63 (10,06 %) sont identiques au signe abrégatif équivalent anglais du fait que les formes développées dans les deux langues comportent des mots qui ont les mêmes initiales. On peut donc considérer que 15 % environ des signes abrégatifs sont français. Un nombre non négligeable de signes abrégatifs sont identiques à l'anglais parce que leur forme développée est calquée sur l'anglais : ainsi, PKC ou ATP sont des sigles très courants, de même que leurs formes développées, mais leur structure syntaxique, caractérisée par une inversion de la tête du syntagme et

de son extension, est empruntée à l'anglais : *protéine kinase C*, *adénosine triphosphate*. On remarquera cependant que la graphie a été francisée (signes diacritiques et *e* muet). La dernière catégorie de sigles, empruntés à l'anglais, constitue la partie la plus importante : 490 signes, soit 75 %, se subdivisent en deux sous-groupes : pour 157 d'entre eux (25,07 %), il existe une forme développée française canonique, mais elle ne donne pas lieu à la production de signes abrégatifs francisés. Pour les 333 restant (53,19 %), il n'a pas été possible de trouver ni dans le corpus, ni dans d'autres sources, de forme développée autre que la forme anglaise, laquelle est généralement introduite entre parenthèses et en italique.

Le phénomène d'emprunt ou de non-francisation peut paraître excessivement important : il représente plus de la moitié des signes abrégatifs. Un examen détaillé révèle que les signes abrégatifs non francisés ont fréquemment un statut de symbole établi ou potentiel : en effet, des comités de nomenclatures utilisant l'anglais comme langue de communication réglementent la dénomination des noms de gènes, de protéines, de microorganismes, des substances chimiques, etc. et en impose l'usage international. Il est donc pratiquement impossible que toute nouvelle molécule découverte soit désignée en contexte autrement qu'à l'aide d'un sigle ayant valeur de symbole translinguistique, et découlant d'une forme développée anglaise. Pour des raisons de commodité, l'usage du signe abrégatif en contexte est alors prédominant. Enfin, seuls certains noms se francisent facilement, lorsqu'il ne s'agit que d'ajouter un *e* muet ou un accent (*culline*), tandis que les autres, trop complexes, restent en l'état (*ADP-ribosylation factor – GTPase activating protein*).

On peut faire l'hypothèse de l'intervention de plusieurs facteurs motivant la raison de la francisation ou son absence :

- la classe référentielle : du fait de l'intervention des nomenclatures internationales, ce sont pour les molécules que les noms anglais semblent être empruntés ;
- la morphologie de la forme développée et du signe abrégatif lui-même : il semble que lorsque les modifications morphologiques sont mineures, la francisation de la forme développée est facilitée (ajout d'un *e* muet, ou d'un accent : *cullin/culline*) ;
- l'ancienneté de la dénomination : la prolifération rapide des noms de molécules nouvelles semble prévenir la francisation des formes développées.

En raison du choix des textes délimitant notre corpus, nous proposons ces explications à titre d'hypothèses explicatives, lesquelles mériteraient d'être vérifiées à l'aide d'une étude *ad hoc* sur un corpus plus important.

7. Introduction en contexte

La manière d'introduire les signes abrégatifs et leurs éventuelles formes développées varient en fonction de la typologie des équivalences décrites ci-dessus. Dans la grande majorité des cas, les règles du français sont respectées, les fluctuations témoignant de la complexité et de la variabilité des phénomènes rencontrés.

7.1. Signes abrégatifs en français avec forme développée en français

7.1.1. La forme développée précède le signe indiqué entre parenthèses à la première occurrence. Le signe est repris, par la suite, seul dans le texte. Cette forme d'introduction des signes abrégatifs ainsi que leur reprise ultérieure est canonique en français :

(22) *Jusqu'à récemment, la cascade d'activation des récepteurs couplés aux protéines G (RCPG) semblait être linéaire et n'impliquer que les protéines G hétérotrimériques. [...] Ainsi, les kinases spécifiques aux RCPG et les arrestines, impliquées dans leur découplage des protéines G, semblent [...].* (Bouvier et al. 2000 : 644)

7.1.2. Le signe abrégatif est accompagné d'une paraphrase, plutôt que de sa forme développée. Si le signe n'est pas très connu, il est cependant préférable de l'accompagner de sa forme développée complète.

(23) *CD4 sert également de co-récepteur pour le virus responsable du SIDA chez l'homme (VIH) lors de l'étape de reconnaissance des cellules cibles.* (Cosson et Letourneur 2000 : 640)

7.2. Signes abrégatifs anglais avec forme développée en français et/ou en anglais

7.2.1. En l'absence d'équivalent français, le signe abrégatif précède la forme développée anglaise qui est présentée en italique et entre parenthèses. Cette façon de procéder est canonique, et l'absence éventuelle de l'italique serait fautive.

(24) *Parmi ces protéines, on retrouve la BiP (immunoglobulin binding protein, Grp78), la Grp94 (glucose-related protein 94 kDa) et la PDI (protein disulfide isomerase).* (Lancôt et Crine 1993 : 1249)

Il est préférable d'éviter d'utiliser la forme développée directement dans la phrase (*[...]codant pour le récepteur de l'insulin-like growth factor-II [IGF-II] humain*) et d'adopter la présentation suivante : *[...] de l'IGF-II humain* (human insulin-like growth factor-II).

7.2.2. L'équivalent français de la forme développée anglaise précède le signe abrégé et la forme développée anglaise en italique, tous deux entre parenthèses. Cette façon de procéder est canonique, l'ordre des éléments dans la parenthèse est sans importance. Il arrive que la forme développée française finisse par donner lieu à un signe abrégé français motivé, si toutefois le signe abrégé anglais ne provient pas d'une nomenclature. Cette méthode représente une bonne façon d'introduire des formulations équivalentes dans la langue d'arrivée.

(25) *Les protéines nucléaires, synthétisées dans le cytoplasme comme toutes les protéines, sont adressées au noyau grâce à un signal de localisation nucléaire (NLS, nuclear localisation signal) constitué d'une courte séquence d'acides aminés basiques, parfois bipartite.* (Schreiber, de Murcia et Ménissier-de Murcia 1992 : 134)

7.2.3. Lorsqu'il n'y a pas d'équivalent français pour la forme développée, une paraphrase explicative, à valeur générique, peut précéder une parenthèse contenant la forme développée anglaise en italique et le signe :

(26) *Elle est ensuite transférée, en formant une autre liaison thioester, sur un membre d'une famille d'enzymes de conjugaison (ubiquitin-conjugating enzymes ou Ubc) encore appelées E2.* (Coux et Piechaczyk 2000 : 623)

7.2.4. Le sigle peut être employé seul dans le texte sans que la forme développée ne soit présentée. Certains signes abrégés sont en effet très connus dans le cercle des scientifiques et n'ont pas besoin d'être définis. La situation pourrait être différente dans un texte de vulgarisation.

(27) *Le grand récepteur présent à la surface cellulaire participe à l'endocytose de l'IGF-II, des hydrolases et, éventuellement, à celle d'autres ligands phosphorylés tels que le TGF et la proliférine.* (Méresse et al. 1993 : 152)

(28) *À l'opposé de l'ADN, qui est stable du point de vue métabolique, les ARN et les protéines sont renouvelés de façon permanente dans la cellule.* (Carillo et al. 1995a : 723)

7.3. Fluctuations

La forme développée, anglaise ou française, peut parfois être présentée seulement à la deuxième ou troisième mention du signe abrégé. À l'inverse, la forme développée peut être répétée plusieurs fois avant l'introduction du signe. Dans les deux cas, il est préférable de présenter la forme développée dès la première occurrence. La correction peut être introduite dans la traduction sans difficulté.

8. Recommandations pour la traduction

Le corpus examiné nous a permis de mettre en évidence un ensemble de comportements des signes abrégatifs utilisés en biologie cellulaire et moléculaire, certains reflétant l'usage habituel des signes, d'autres relevant spécifiquement du domaine, et quelques-uns reflétant la fluctuation dans l'usage. À la lumière de l'analyse réalisée, nous proposons quelques pistes facilitant le processus de traduction.

8.1. Recherche documentaire

La première étape consiste à distinguer les symboles courants des autres signes abrégatifs. Les règles à suivre pourront être consultées dans des ouvrages de référence de la langue française, sachant que parfois les exigences éditoriales pourront prendre le pas sur certaines conventions. Pour les autres signes abrégatifs, il faut en premier lieu rechercher la présence d'une forme développée, ce qui permettra, du même coup, d'accéder au référent et à la catégorie référentielle concernée. Si elle n'est pas présente, il faudra chercher des indices contextuels, et éventuellement faire des vérifications à l'aide de Medline, par exemple afin d'éviter une erreur liée à une éventuelle homonymie. À ce stade, il faut distinguer les signes donnant un accès direct au référent, et ceux, employés en apposition, qui ont une fonction relationnelle. Il est alors possible de consulter des sources spécifiques, en fonction du type de référent : pour les gènes, molécules diverses, les protéines, les enzymes, toute substance ayant un rôle biologique, on se dirigera vers la consultation en ligne des bases de données *ad hoc*, de même que pour les substances chimiques, les souches d'animaux de laboratoire et les lignées cellulaires. Pour le reste, les ouvrages listant les abréviations rencontrées en médecine ou en biologie ainsi que leurs équivalents seront précieuses. Enfin, il faudra consulter des textes rédigés en français, les sources parmi les plus utiles à ce chapitre étant les articles de *m/s médecine sciences* en ligne et la base de données de l'INIST. Il restera à compiler les informations pour passer à la phase de la décision.

8.2. Décisions de traduction

8.2.1. Francisation

Lorsque le signe abrégatif est un symbole dénominatif ou renvoie à une dénomination commune normalisée (également repérable dans les nomenclatures de produits chimiques ou de médicaments), il faut impérativement respecter les normes imposées, c'est-à-dire suivre les nomenclatures. Il faut suivre la

graphie indiquée dans ces dernières, mais on ne remplacera pas systématiquement un sigle indiqué comme secondaire par le sigle synonyme consensuel : ce peut être un choix de l'auteur. La plupart du temps, dans les autres cas, l'abréviation est francisée.

8.2.2. Introduction des formes développées

Les ouvrages de références de la langue française (par ex. De Villers 2001 ; Bureau de la traduction 1996 ; Grevisse 1993), il est conseillé d'inscrire la forme développée d'un sigle ou d'un acronyme au long dès sa première apparition dans le texte. Mortureux (1994 : 18) précise que « [l'] usage est de juxtaposer le terme abrégé et son développement, dans un ordre variable, le second des deux étant, suivant le cas, mis entre parenthèses ou traité comme une incise, entre virgules ». Il est à noter également que, dans les articles du corpus, la forme développée de certains sigles a été retrouvée dans la légende des figures seulement, pour en faciliter leur repérage, et non dans le texte. Selon le cas, on introduira la forme développée française, ou anglaise, avec la forme développée pertinente (voir plus haut).

8.2.3. Genre et déterminant

La règle stipule que « [c]ontrairement à l'usage anglais, les sigles s'emploient généralement avec l'article » (Bureau de la traduction 1996 : 14). De plus, l'article d'un sigle anglais sera du genre de l'équivalent français de la tête du syntagme (ou déterminé) de la forme développée. Le non-emploi de l'article devant les sigles leur donne un caractère de nom propre (Kleiber, 1996 ; Grevisse 1993 : § 451). De façon générale, les noms de molécule se comportent comme des noms propres, mais il ne faut pas étendre cette règle à l'ensemble des substances chimiques. Nous avons montré, avec les différents cas de figure, qu'il fallait traiter le genre et le déterminant selon le type de signe abrégé rencontré.

8.2.4. Pluriel des sigles

Le guide du rédacteur (1996 : 14) stipule que « les sigles, d'origine française ou étrangère, restent invariables tant qu'ils ne sont pas clairement lexicalisés ». En dépit de quelques écarts rencontrés dans le corpus (*les ITAMs*, *les t-SNAREs*), cette règle doit être respectée. Il s'agit, en fait, d'un calque grammatical de l'anglais

9. Conclusion

En conclusion, il est remarquable que la biologie cellulaire et moléculaire fasse usage de si nombreux signes abrégés, au comportement si versatile.

C'est une des particularités de ce domaine. Ces signes abrégatifs ne suivent pas les règles canoniques d'équivalence et de comportement discursif. Il serait intéressant de savoir si, avec leur prolifération, les règles d'usage se stabilisent ou si au contraire les fluctuations augmentent. Avec l'essor des nomenclatures de toute sorte maintenant en accès libre sur Internet, on peut se demander si celles-ci participent à une éventuelle stabilisation, laquelle serait alors liée à une intégration, chez les usagers, de la nature symbolique translinguistique des sigles à fonction dénominative.

Remerciements

Manon Pageau a été récipiendaire d'une bourse de recherche octroyée par le FCAR (Fonds de recherche sur la nature et les technologies, Québec) pour 2001 et 2002. Une partie du travail a été rendue possible grâce à une subvention de recherche octroyée à Sylvie Vandaele par le Conseil de recherche en sciences humaines du Canada. Nous remercions Mariana Raffo pour la collecte de données additionnelles.

Bibliographie

- BERGERON (M.) et Kahn (A.) 1995, « m/s médecine/sciences : À 10 ans, la médecine et la science dans dix ans », in *m/s médecine/sciences*, Montréal-Paris, 11/3, pp. 315-316.
- Bureau de la traduction 1996, *Le guide du rédacteur*, Ottawa.
- CALVET (L.-J.) 1970, *Les sigles*, Paris, P.U.F.
- COLLET (T.) 2000, *La réduction des unités terminologiques complexes de type syntagmatique*, Thèse de doctorat, Montréal, Université de Montréal.
- DE VILLERS (M.-E.) 2001, *Multidictionnaire de la langue française version électronique*, Montréal, Québec Amérique, version 1.0.
- DIONNE (S.) et BOULAY (J.) 1998, *Guide bilingue des abréviations médicales*, 3^e éd., Montréal, MacLean Hunter Santé.
- GERMAIN (L.) et LAPIERRE (A.) 1988, « Le sigle. Définition, caractéristiques et emploi », in *Cahiers de lexicologie*, 52/2, pp. 55-74.
- GLADSTONE (W.J.) 1990, *Dictionnaire anglais-français des sciences médicales et paramédicales*, 3^e éd., Saint-Hyacinthe-Paris, Edisem-Maloine.
- GREVISSE (M.) 1993, *Le bon usage : grammaire française*, 13^e éd., Paris-Gembloux, Duculot.

- JABLONSKI (J.) 2005, *Dictionary of medical acronyms & abbreviations*, Philadelphia, Saunders.
- JAMMAL (A.) 1992, « L'hybridation du langage médical français un phénomène inévitable ? », in *Journal de radiologie*, 73/3, pp. 213-214.
- KOCOUREK (R.) 1982, *La langue française de la technique et de la science*, Paris, La documentation française.
- KLEIBER (G.) 1996, *Noms propres et noms communs : un problème de dénomination*, in *Meta*, XLI/4, pp. 567-589.
- MAILLET (M.) 1996, *Dictionnaire des abréviations en biologie : 2500 abréviations et acronymes*, Paris, Masson.
- MORTUREUX (M.-F.) 1994, « Siglaison-acronymie et néologie lexicale », in *Les sigles*, Nanterre, Centre de Recherches Linguistiques, pp. 11-32.
- MONNIER (P.) 1994, « Usages et formations des sigles : Une application dans l'industrie spatiale », in *Les sigles*, Nanterre, Centre de Recherches Linguistiques, pp. 91-108.
- PUSTEJOVSKY (J.), CASTAÑO (J.), SAURÍ (R.), RUMSHISKY (A.), ZHANG (J.) et LUO (W.) 2002, « Medstract : Creating Large-scale Information Servers for Biomedical Libraries », in *ACL 2002 Workshop on Natural Language Processing in the Biomedical Domain*, Philadelphia, PA.
- PAGEAU (M.) 2003, *Études descriptives des emprunts à l'anglais dans le domaine de la biologie cellulaire*, Montréal, Université de Montréal.
- Le Petit Robert. Dictionnaire de la langue française* 2001, Paris, Dictionnaires Le Robert, cédérom.
- TOUATI (M.A.) 1994, *Dictionnaire international des abréviations médicales*, Paris, La maison du dictionnaire.
- VANDAELE (S.) 2001a, « Utilisation des bases de données bibliographiques spécialisées », in *Meta*, 46/1, pp. 103-116.
- VANDAELE (S.) 2001b, « Bibliographie sélective pour la traduction des sciences biomédicales », in *Meta*, 46/1, pp. 118-127.
- VANDAELE (S.) et LUBIN (L.) 2005, « Approche cognitive de la traduction dans les langues de spécialité : vers une systématisation de la conceptualisation métaphorique », in *Meta*, 50/2, pp. 337-794.
- WAIN (H.M.) *et al.* 2002, « Guidelines for Human Gene Nomenclature », in *Genomics*, 79/4, pp. 464-470.

Bibliographie des exemples

- AILHAUD (G.) 1998, « L'adipocyte, cellule sécrétrice et endocrine », in *m/s médecine/sciences*, 14/8-9, pp. 858-864.
- BOUVIER (M.) *et al.*, 2000, « Échafaudages protéiques et signalisation modulaire via les récepteurs à sept domaines transmembranaires : au-delà des protéines G », in *m/s médecine/sciences*, 16/5, pp. 644-651.
- BROUTIN (I.) et DUCRUIX (A.) 2000, « Domaines structuraux et signalisation », in *m/s médecine/sciences*, 16/5, pp. 611-616.
- CARILLO (S.) *et al.* 1995, « Le catabolisme protéique intracellulaire : une fonction biologique majeure. Partie I : les mécanismes de dégradation », in *m/s médecine/sciences*, 11/5, pp. 723-734.
- CHARRON (J.) et JEANNOTTE (L.) 1999, « Le rôle essentiel de MEK1 lors de l'angiogenèse placentaire », in *m/s médecine/sciences*, 15/10, pp. 1155-1157.
- COSSON (P.) et LETOURNEUR (F.) 2000, « Transport membranaire dans les cellules eucaryotes », in *m/s médecine/sciences*, 16/5, pp. 635-643.
- COUX (O.) et PIECHACZYK (M.) 2000, « Le système ubiquitine/protéasome : un ensemble (de) complexe(s) pour dégrader les protéines », in *m/s médecine/sciences*, 16/5, pp. 623-629.
- DORSEUIL (O.) 1998, « Petite protéine G Ran et contrôle de l'import-export nucléaire », in *m/s médecine/sciences*, 14/1, pp. 85-89.
- GISSELBRECHT (S.) 1999, « Les protéines de la famille CIS-SOCS : des modulateurs des effets biologiques des cytokines », in *m/s médecine/sciences*, 15/11, pp. 1256-1267.
- GOUD (B.) 1995, « Le voyage intracellulaire des protéines », in *La Recherche*, 274/26, pp. 254-260.
- GOUD (B.) 1999, « Le code d'adressage des protéines », in *m/s médecine/sciences*, 15/11, pp. 1336-1338.
- LANCTÔT (C.) et CRINE (P.) 1993, « La rétention des protéines dans le réticulum endoplasmique », in *m/s médecine/sciences*, 9/11, pp. 1249-1251.
- LECOQ-LAFON (C.), LACOMBE (C.) et MAYEUX (P.) 1999, « GAB1 : un adaptateur protéique multifonctionnel », in *m/s médecine/sciences*, 15/1, pp. 69-72.
- LESCA (P.) et PINEAU (T.) 1999, « Toxicité de la dioxine : rôles des protéines PAS (étapes de transmission du signal) », in *m/s médecine/sciences*, 15/12, pp. 1379-1387.
- MAUVIEL (A.) 1999, « Récentes avancées dans la compréhension de la voie de signalisation du TGF- β par les Smad », in *m/s médecine/sciences*, 15/4, pp. 1535-1537.

- MÉRESSE (S.) *et al.* 1993, « Bases moléculaires du transport vers les lysosomes », in *m/s médecine/sciences*, 9/2, pp. 148-156.
- MONTIXI (C.) et HE (H.-T.) 1999, « Radeaux lipidiques : rôle dans l'activation lymphocytaire », in *m/s médecine/sciences*, 15/10, pp. 1142-1147.
- PAULOIN (A.) 1993, « Bréfeldine A, protéines G et transports membranaires golgiens », in *m/s médecine/sciences*, 9/8-9, pp. 917-925.
- PELENGARIS (S.A.) et EVANS (G.I.) 1997, « Matters of Life and Cell Death », in *Review of Neurology*, 153, pp. 470-477.
- POULIN (F.) et PYRONNET (S.) 2000, « Interactions moléculaires et initiation de la synthèse protéique », in *m/s médecine/sciences*, 16/5, pp. 617-622.
- ROUSSEAU (D.) 2001, « eIF-4E, régulation de la traduction et progression tumorale », in *m/s médecine/sciences*, 17/3, pp. 336-343.
- SCHREIBER (V.), DE MURCIA (G.) et MÉNISSIER-DE MURCIA (J.) 1992, « Le transport des protéines dans le noyau : les signaux de localisation nucléaire et leurs récepteurs », in *m/s médecine/sciences*, 8/2, pp. 134-139.